Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Development and Validation of an Elastic Net–Softmax Machine Learning Model for Predicting Lymph Node Regression Grade After Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer

Kai Liu
Affiliated Cancer Hospital of Xinjiang Medical University, Xinjiang Uygur Autonomous Region
Shuyuan Zhao
Affiliated Cancer Hospital of Xinjiang Medical University
Ran Zhang
Affiliated Cancer Hospital of Xinjiang Medical University, Xinjiang Uygur Autonomous Region
Yan Liu
m18290886651_1@163.com
Affiliated Cancer Hospital of Xinjiang Medical University, Xinjiang Uygur Autonomous Region

Research Article

Keywords: locally advanced rectal cancer, neoadjuvant chemoradiotherapy, lymph node regression grade, Elastic Net, Softmax, machine learning, predictive model

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10882836/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Development and Validation of an Elastic Net–Softmax Machine Learning Model for Predicting Lymph Node Regression Grade After Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer

Authors: [Kai Liu][1], [Shuyuan Zhao][1], [Ran Zhang][2], [Yan Liu]*

Affiliations: [1, 1, 2]Department of Thoracic and Abdominal Radiation Oncology , Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi, Xinjiang, China

* Corresponding author: [Yan Liu], Department of [Thoracic and Abdominal Radiation Oncology], Affiliated Cancer Hospital of Xinjiang Medical University, [No. 789, Suzhou East Street, Urumqi, Xinjiang Uygur Autonomous Region, 830000, China]. Tel: [18290886651]; Email: [m18290886651_1@163.com]

HIGHLIGHTS

An Elastic Net–Softmax model was developed to predict six-grade LRGmax after nCRT. Validation AUC was 0.86 with high specificity (0.87) for low-risk identification. LRGmax correlated with radiotherapy response and 3-year survival, not metastasis. The model is a proof-of-concept tool for preoperative nodal response assessment.

ABSTRACT

Background: Neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision is standard for locally advanced rectal cancer (LARC). While tumor regression grade is routinely reported, lymph node regression grade (LRG) is underutilized. We aimed to develop a machine learning model for preoperative LRG prediction.

Methods: We retrospectively reviewed 111 LARC patients who underwent nCRT plus surgery between January 2016 and December 2020. LRGmax (grades 0–5) was assessed using Mandard criteria. Elastic Net regression with 10-fold cross-validation was used for feature selection, followed by a Softmax multinomial classifier. Patients were randomly split into training (70%) and validation (30%) sets. Performance was evaluated using AUC, accuracy, sensitivity, specificity, precision, and recall. Bootstrap resampling and decision curve analysis assessed calibration and clinical net benefit.

Results: The LRGmax distribution was: grade 0 (28.8%), 1 (39.6%), 2 (12.6%), 3 (10.8%), 4 (5.4%), and 5 (2.7%). In the training set, AUC was 0.99. In the validation set, AUC was 0.86 and specificity 0.87. Calibration curves showed high consistency; DCA revealed positive net benefit. Higher LRGmax correlated with increased nodal metastasis count, lymph node ratio, and worse survival (all $P < 0.001$). LRGmax correlated positively with radiotherapy efficacy ($r = 0.28$, $P = 0.003$) and negatively with 3-year survival ($r = –0.47$, $P < 0.001$).

Conclusions: The Elastic Net–Softmax model showed discriminative utility for predicting LRGmax after nCRT in LARC. Despite decreased validation metrics due to limited sample size and class imbalance, the model retained predictive value as a proof-of-concept tool. Multi-center prospective studies are warranted.

Keywords: locally advanced rectal cancer; neoadjuvant chemoradiotherapy; lymph node regression grade; Elastic Net; Softmax; machine learning; predictive model

1. Introduction

Colorectal cancer is the third most common malignancy and the second leading cause of cancer-related mortality worldwide [1,2]. In China, rectal cancer predominantly presents as mid-to-low lesions, comprising over 80% of all cases [3]. Neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision (TME) is the standard therapeutic paradigm for locally advanced rectal cancer (LARC) [4,5]. Clinical response assessment, typically performed 6–8 weeks after nCRT completion, directly influences subsequent treatment decisions.

Tumor regression grade (TRG) is a well-established pathological metric for evaluating primary tumor response following nCRT [6]. Analogously, lymph node regression grade (LRG) assesses the treatment response of metastatic lymph nodes based on postoperative pathology [7,8]. However, while TRG has been incorporated into routine pathology reports, LRG evaluation has not

received adequate attention. Lymph node metastasis status represents one of the most critical prognostic factors in rectal cancer, and the number of metastatic lymph nodes currently serves as the sole basis for ypN staging. Persiani et al. [9] demonstrated that neoadjuvant therapy reduces both the total number of lymph nodes retrieved and the number of positive nodes, thereby compromising the accuracy of ypN staging. Consequently, evaluating TRG without considering LRG provides an incomplete picture of the overall treatment response.

The present study retrospectively analyzed clinical data from LARC patients who received preoperative nCRT. By integrating clinical, imaging, and pathological variables, we employed machine learning methodologies to construct a novel LRG prediction model, aiming to comprehensively assess nCRT efficacy and provide data support for individualized treatment decision-making.

2. Materials and Methods

2.1 Patients

We retrospectively reviewed patients who received nCRT followed by surgery at the Affiliated Cancer Hospital of Xinjiang Medical University between January 2016 and December 2020.

Inclusion criteria: (i) age 18–80 years; (ii) Karnofsky Performance Status (KPS) score ≥ 70; (iii) histologically confirmed rectal adenocarcinoma; (iv) clinically staged as locally advanced (T3–4/N+) mid-to-low rectal cancer (tumor distance from anal verge ≤ 12 cm) by pre-treatment imaging; (v) no prior pelvic radiotherapy or malignancy; (vi) no severe cardiopulmonary, hepatic, or renal comorbidities; and (vii) completion of nCRT followed by radical TME.

Exclusion criteria: (i) treatment interruption > 2 weeks; (ii) distant metastasis on imaging; or (iii) incomplete medical records.

This study was approved by the Ethics Committee of the Affiliated Cancer Hospital of Xinjiang Medical University (approval number: [XXX]). Written informed consent was obtained from all patients. The study was conducted in accordance with the Declaration of Helsinki.

2.2 Treatment Protocols

Radiotherapy. All patients underwent intensity-modulated radiation therapy (IMRT) with a thermoplastic immobilization device. Simulation CT was performed (slice thickness 5 mm, from L3 to 10 cm below the ischial tuberosity). Target volumes were delineated according to the RTOG consensus: gross tumor volume (GTV) included visible primary tumor and metastatic lymph nodes; clinical target volume (CTV) encompassed GTV plus corresponding nodal drainage regions; planning target volume (PTV) was generated by expanding CTV with a 0.5–1.0 cm margin. A total dose of 45–50.4 Gy in 25–28 fractions was delivered to 95% of the PTV using 6-MV X-rays via 5- or 7-field IMRT. Dose constraints for organs at risk were: femoral head V20 < 50% and V50 < 5%; bladder V50 < 50%; and small bowel V50 < 5%.

Concurrent chemotherapy. All patients received concurrent chemotherapy, either: (i) capecitabine 850 mg/m² orally twice daily on radiotherapy days; or (ii) XELOX: oxaliplatin 130 mg/m² intravenously on day 1 plus capecitabine 1,000 mg/m² orally twice daily on days 1–14, every 21 days.

Surgery. Radical resection was performed 6–8 weeks after nCRT completion, adhering to TME and R0 resection principles.

2.3 Pathological Assessment

TRG and LRG were evaluated independently by two experienced pathologists in a blinded manner, with discrepancies resolved by a senior pathologist.

TRG was assessed using the Dworak/Rodel criteria: TRG 4, complete regression (no viable tumor cells, only fibrosis); TRG 3, marked regression (> 50% tumor cells with degenerative changes and predominant fibrosis); TRG 2, moderate regression (26–50% tumor regression with obvious fibrosis); TRG 1, minimal regression (< 25% tumor regression with obvious fibrosis); and TRG 0, no regression (Figure 1).

LRG was assessed using Mandard's criteria (1994): LRG 0, normal lymph node without tumor cells; LRG 1, 100% fibrosis with no residual tumor cells; LRG 2, 75–100% fibrosis with 0–25% tumor cells; LRG 3, 50–75% fibrosis with 25–50% tumor cells; LRG 4, 25–50% fibrosis with 50–75% tumor cells; and LRG 5, 0–25% fibrosis with 75–100% tumor cells (Figure 2).

Because nodal response varies across individual lymph nodes, three composite metrics were calculated: **LRGmax** (highest score among all nodes), **LRGsum** (arithmetic sum of all scores), and **lymph node ratio (LNR)** (metastatic nodes divided by total nodes harvested).

2.4 Statistical Analysis

Statistical analyses were performed using SPSS 27.0 (IBM Corp., Armonk, NY, USA) and Python 3.11. Normality of continuous variables was assessed by the Shapiro–Wilk test. Normally distributed data were compared by independent-samples *t*-test or one-way ANOVA, and non-normally distributed data by Mann–Whitney *U* test or Kruskal–Wallis *H* test. Categorical variables were compared by χ^2 test or Fisher's exact test. Associations between LRG and clinical variables were evaluated by Spearman's rank correlation. Two-sided tests were used, and $P < 0.05$ indicated statistical significance.

2.5 Model Development and Evaluation

The LRG prediction model was developed in two stages. Feature selection was performed by Elastic Net regression with 10-fold cross-validation. A multinomial Softmax classifier was then trained. Patients were randomly split into training and validation sets (7:3). Discrimination was assessed by ROC curves and AUC. Calibration was evaluated by bootstrap resampling (1,000 iterations). Clinical utility was assessed by DCA. Softmax probability heatmaps visualized classification performance.

For multiclass evaluation, macro-averaged one-vs-rest (OvR) AUC was calculated. Accuracy, precision, recall, and specificity were macro-averaged across all six LRGmax classes. Therapeutic efficacy was ordinally coded (CR = 1, PR = 2, SD = 3, PD = 4; higher values indicated poorer response) for correlation analysis.

Elastic Net retained 12 variables, including radiotherapy response, TRG, pathological stage, EMVI, chemotherapy regimen, and CA724 change. Neither synthetic oversampling nor class-weight adjustment was applied due to limited sample size, which is acknowledged as a limitation.

3. Results

3.1 Baseline Characteristics

A total of 111 LARC patients were enrolled. According to postoperative LRGmax, the distribution was: grade 0 (32 cases, 28.8%), grade 1 (44, 39.6%), grade 2 (14, 12.6%), grade 3 (12, 10.8%), grade 4 (6, 5.4%), and grade 5 (3, 2.7%). No significant differences were observed among groups in age, sex, body surface area, body weight, CEA, CA724, clinical T/N stage, MRF status,

comorbidities, family history, or pathological differentiation (all $P > 0.05$), indicating good comparability.

Significant differences were found in post-treatment CA724 change ($P = 0.027$), KPS score ($P = 0.007$), EMVI status ($P = 0.037$), radiotherapy efficacy ($P < 0.001$), pT stage ($P = 0.029$), pN stage ($P < 0.001$), TRG ($P < 0.001$), chemotherapy regimen ($P = 0.042$), and pathological stage ($P < 0.001$). As LRGmax grade increased, lymph node metastasis count and LNR progressively increased (both $P < 0.001$), while overall survival and 3-year overall survival rate decreased (both $P < 0.001$) (Table 1).

3.3 Feature Selection and Model Construction

Elastic Net regression was applied for feature selection. As the penalty parameter λ increased, coefficients of irrelevant variables shrank to zero, retaining only key predictors associated with LRG classification. Ten-fold cross-validation determined the optimal λ at the minimum cross-validation error, upon which the final Softmax model was established (Figure 3).

The Softmax model coefficient plot reveals the transparency characteristics of the model's decision-making process from an interpretability perspective: radiological response, primary lesion TRG, pathological stage, and EMVI status constitute the core set of features for LRGmax grading prediction; meanwhile, the coefficients for certain traditional prognostic indicators (such as OS and KPS) were compressed to values close to zero by Elastic Net regularization, demonstrating the model's ability to automatically select relevant features and its discriminative stability within a complex clinical feature space (Figure 4).

3.4 Classification Performance and Prediction Confidence

The Kruskal–Wallis test demonstrated significant differences in LNR and metastatic lymph node counts across both TRG and LRGmax grades (all $P < 0.001$). As LRGmax increased, LNR and the number of metastatic lymph nodes showed rising trends, indicating progressive nodal metastatic burden with poorer nodal regression (Figure 5).

The Softmax probability heatmap revealed that most patients had their LRGmax grade assigned the highest predicted probability, with only minority cases exhibiting cross-predictions between adjacent categories, indicating high classification consistency. Minor uncertainty was observed for LRGmax 3, where some samples displayed probabilities of 0.4–0.6 in neighboring columns, suggesting feature overlap at classification boundaries. Despite limited sample

sizes, LRGmax 4 and 5 maintained concentrated predictive distributions (Figure 6).

3.5 Discriminative Ability Evaluation

In the training set, the model achieved a macro-averaged AUC of 0.99, accuracy of 0.88, precision/recall/sensitivity of 0.94, and specificity of 0.97. In the validation set, the AUC was 0.86, with specificity of 0.87. Although validation-set accuracy decreased to 0.43 and sensitivity to 0.30, the retained high specificity (0.87) indicated reliable identification of low-risk patients (LRGmax 0–1) (Table 3, Figure 7).

3.6 Calibration and Clinical Utility

Calibration curves showed that the bias-corrected curve closely overlapped with the ideal reference line after bootstrap validation, indicating good calibration (Figure 8). DCA demonstrated that this model exhibits high reliability in identifying LRGmax = 1, but has limitations when handling LRGmax 5; therefore, in practical applications, it is necessary to implement differentiated threshold strategies tailored to different categories (Figure 9).

3.7 Correlation Between LRG and Clinical Outcomes

LRGmax was significantly associated with poorer radiotherapy response (Spearman $r = 0.28$, $P = 0.003$) and negatively correlated with 3-year overall survival ($r = –0.47$, $P < 0.001$). TRG showed a stronger association with radiotherapy response ($\chi^2 = 82.03$, $P < 0.001$) and was also associated with 3-year survival ($\chi^2 = 26.15$, $P < 0.001$). Neither LRGmax nor TRG correlated significantly with distant metastasis (both $P > 0.05$), suggesting that LRGmax may better reflect long-term prognosis than distant dissemination risk. This indicates that LRGmax may serve as a potential prognostic biomarker integrating imaging-based features with tumor burden information (Table 4, Figure 10).

4. Discussion

Neoadjuvant chemoradiotherapy (nCRT) combined with total mesorectal ex cision (TME) remains the standard of care for locally advanced rectal can cer (LARC). Yet primary and nodal regression do not always parallel each other: some patients show substantial primary tumor response while still harboring residual nodal metastases, which compromises long-term progno

sis [3,10]. The OPRA trial further showed that baseline nodal involvement was independently associated with organ-preservation failure (HR 1.89) [11], highlighting the need for accurate preoperative assessment of nodal response. A predictive model for lymph node regression grade (LRG) may therefore improve preoperative risk stratification [12].

We chose Elastic Net regression for feature selection and a Softmax multinomial classifier for LRGmax prediction. Elastic Net combines LASSO and Ridge penalties and offers better stability than either alone in high-dimensional, small-sample settings [13]. Softmax, meanwhile, allows simultaneous multiclass probability estimation, which suits the graded nature of nodal response [14]. LRGmax is fundamentally an ordinal variable, ranging from complete regression (grade 0) to no regression (grade 5). Ordinal models such as rank-consistent neural networks could in theory exploit this ordering to sharpen boundary discrimination [15]. However, recent work indicates that the advantage of ordinal over nominal classifiers depends heavily on dataset and task characteristics [15]. In our cohort, severe class imbalance—particularly for grades 4 and 5 (n = 6 and n = 3)—raised concerns about sparse-data bias and violation of the proportional odds assumption. We therefore favored Softmax for its flexibility, while acknowledging that future large-scale studies should directly compare ordinal and multinomial frameworks using rank-consistent metrics .

The model achieved near-perfect discrimination in the training set (AUC 0.99) but dropped markedly in the validation set (AUC 0.86), with accuracy, sensitivity, and precision falling to 0.43, 0.30, and 0.30, respectively. This decline likely stems from three sources. First, the single-center cohort comprised only 111 patients, and grades 4 and 5 accounted for merely 5.4% and 2.7% of cases, leaving too few examples for the model to learn minority-class patterns. Second, six-grade classification is intrinsically difficult because adjacent grades represent continuous biological transitions along the pathological response spectrum; feature spaces therefore overlap, producing boundary misclassification, as seen in the off-diagonal signals of the probability heatmap. Third, the 7:3 split yielded a validation set of just 32 patients, magnifying sampling variability. The absence of resampling techniques (e.g., SMOTE) or cost-sensitive learning further constrained performance. A recent review showed that hybrid strategies combining oversampling with ensemble or class-weighted algorithms consistently outperform single-method approaches in imbalanced clinical data [16], pointing to a practical direction for refinement.

Despite these constraints, the model retains useful information. The validation AUC of 0.86 indicates discrimination well above chance; high specificity (0.87) allows reliable identification of low-risk patients (LRGmax 0–1), which matters for screening; calibration curves support trustworthy probability estimates; and decision curve analysis shows positive net benefit across a broad threshold range. These observations suggest that predicting LRG

max from structured clinical data is feasible, though the current model should serve only as an auxiliary tool rather than a standalone basis for clinical decisions.

The diagonal dominance of the Softmax probability heatmap indicates that the model extracts meaningful features. The dispersion around LRGmax 3—where neighboring columns show probabilities of 0.4–0.6—mirrors the pathological reality of nodal response: grade 3 represents an intermediate state with mixed histological components that share spectral similarity with adjacent grades, a pattern frequently observed in ordinal multiclass tasks [6]. Among the selected variables, post-treatment CA724 change, KPS score, EMVI status, radiotherapy response, TRG, and pathological stage were significantly associated with LRG. EMVI positivity signals strong invasiveness and vascular dissemination potential, consistent with poorer treatment response [17]. Together, these associations indicate that nodal regression reflects tumor burden, host functional status, and treatment sensitivity.

Correlation analyses also support the prognostic relevance of LRG. LRGmax correlated with poorer radiotherapy response and lower 3-year survival, but not with distant metastasis. This implies that LRGmax captures local-regional response and long-term survival risk rather than systemic spread. Li et al. [18] reported that LRG independently predicted disease-free survival, whereas TRG was linked to distant metastasis. Our findings partially agree, suggesting that primary and nodal regression track distinct biological processes; evaluating both TRG and LRG together should therefore provide a more comprehensive assessment of prognosis [3,11].

Clinically, preoperative LRGmax prediction could refine risk stratification for organ-preservation strategies. Current watch-and-wait protocols rely primarily on primary tumor response (TRG or cCR), yet a secondary analysis of the OPRA trial found that even among patients who achieved clinical complete response, baseline nodal involvement and near-complete response status were independently associated with tumor regrowth and organ-preservation failure [19].Residual nodal disease is thus a critical driver of local recurrence that primary tumor assessment alone may miss [12]. A validated LRGmax model could identify patients with a high probability of complete nodal regression as potential candidates for non-operative management, while flagging those with predicted poor nodal response for intensified systemic therapy or extended lateral lymph node dissection [7,10].

This study has several limitations. Its single-center, retrospective design, limited sample size, and marked class imbalance may have impaired recognition of minority classes. Only internal validation was performed, and external validation is still needed. Moreover, the model relied solely on clinical and pathological variables; MRI radiomics and molecular biomarkers were not included. Multimodal approaches have shown clear promise: Feng et al. [20] combined MRI radiomics with pathological features to predict pathological complete response with an AUC of 0.86–0.87, and the PRISM-CRC

framework, which fuses whole-slide pathology, CT imaging, and clinical data, achieved a 5-year disease-free survival C-index of 0.82 [21]. Incorporating imaging and molecular data into future iterations could improve predictive performance.

In conclusion, we developed an Elastic Net–Softmax model for predicting LRGmax after nCRT in LARC. Validation metrics declined because of limited sample size and class imbalance, yet the model preserved meaningful discriminative and clinical utility as an exploratory auxiliary tool. Multi-center, large-sample, prospective studies are needed to optimize and externally validate the model before it can be translated into routine clinical practice.

5. Conclusions

1.The Elastic Net–Softmax model demonstrated favorable discriminative ability, calibration performance, and clinical net benefit for predicting LRGmax after nCRT in LARC.

2.Despite decreased validation metrics attributable to limited sample size and severe class imbalance, the model retained meaningful predictive value, particularly for identifying low-risk patients (LRGmax 0–1).

3.LRGmax correlated with radiotherapy response and 3-year overall survival, supporting its prognostic relevance as a complementary biomarker to TRG.

4.Multi-center, large-sample, prospective studies incorporating multimodal data (radiomics, molecular biomarkers, and deep learning features) are warranted to enhance model generalizability and clinical translation.

6 . REFERENCES

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263.

2. Han B, Zheng R, Zeng H, et al. Cancer incidence and mortality in China, 2022. J Natl Cancer Cent. 2024;4(1):47-53.

3. He L, Xiao J, Zheng P, et al. Lymph node regression grading of locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy. World J Gastrointest Oncol. 2022;14(8):1429-1445.

4. Conroy T, Bosset JF, Etienne PL, et al. Neoadjuvant chemotherapy with FOLFIRINOX and preoperative chemoradiotherapy for patients with locally advanced rectal cancer (UNICANCER-PRODIGE 23): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2021;22(5):702-715.

5. Bahadoer RR, Dijkstra EA, van Etten B, et al. Short-course radiotherapy followed by chemotherapy before total mesorectal excision (TME) versus preoperative chemoradiotherapy, TME, and optional adjuvant chemotherapy in locally advanced rectal cancer (RAPIDO): a randomised, open-label, phase 3 trial. Lancet Oncol. 2021;22(1):29-42.
6. Shin J, Seo N, Kim H, et al. MRI Radiomics Model Predicts Pathologic Complete Response of Rectal Cancer Following Chemoradiotherapy. Radiology. 2022;303(2):351-358.
7. Kroon HM, Hoogervorst LA, van der Wolf FS, et al. Systematic review and meta-analysis of long-term oncological outcomes of lateral lymph node dissection for metastatic nodes after neoadjuvant chemoradiotherapy in rectal cancer. Eur J Surg Oncol. 2022;48(7):1475-1482.
8. Kohl VKB, Weber K, Brunner M, et al. Factors influencing downstaging after neoadjuvant long-course chemoradiotherapy in rectal carcinoma. Int J Colorectal Dis. 2022;37(6):1355-1365.
9. Persiani G, Li Destri G, Maugeri A, et al. The prognostic impact of neoadjuvant chemoradiotherapy on lymph node sampling in patients with locally advanced rectal cancer. Updates Surg. 2020;72(4):1131-1139.
10. O'Connell LV, Murnaghan N, O'Brien A, et al. Significance of nodal yield after rectal resection in the TNT era: does count matter? BMC Cancer. 2026 Mar 2;26(1):446. Garcia-Aguilar J, Patil S, Gollub MJ, et al. Organ Preservation in Patients With Rectal Adenocarcinoma Treated With Total Neoadjuvant Therapy. J Clin Oncol. 2022;40(23):2546-2556.
11. Sun Y, Wu X, Lin H, Lu X, Huang Y, Chi P. Lymph Node Regression to Neoadjuvant Chemoradiotherapy in Patients with Locally Advanced Rectal Cancer: Prognostic Implication and a Predictive Model. J Gastrointest Surg. 2021 Apr;25(4):1019-1028.

12. Abbaspour E, Mansoori B, Karimzadhagh S, Chalian M, Pouramini A, Sheida F, Daskareh M, Haseli S. Machine learning and deep learning models for preoperative detection of lymph node metastasis in colorectal cancer: a systematic review and meta-analysis. Abdom Radiol (NY). 2025 May;50(5):1927-1941.
13. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. 2005;67(2):301-320.
14. Bridle JS. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: Soulie FF, Herault J, eds. Neurocomputing: Algorithms, Architectures and Applications. Berlin: Springer; 1990:227-236.
15. Litvinov A, Ushakov E, Senotrusova S, et al. Clinically aware learning: ordinal loss improves medical image classifiers. J Clin Med. 2026;15(1):365.
16. Salmi M, Atif D, Oliva D, et al. Handling imbalanced medical datasets: review of a decade of research. J Biomed Inform. 2024;57:83-94.
17. Chen S, Li N, Tang Y, et al. The prognostic value of MRI-detected extramural vascular invasion (mrEMVI) for rectal cancer patients treated with neoadjuvant therapy: a meta-analysis. Eur Radiol. 2021;31(12):8827-8837.

18. Li J, Yuan J, Liu H, et al. Lymph nodes regression grade is a predictive marker f or rectal cancer after neoadjuvant therapy and radical surgery. Oncotarget. 2016;7(6): 10375-10384.
19. Zambare W, Miranda J, Horvat N, et al. Assessing the OPRA trial for surgical on cologists: safety and feasibility of a total neoadjuvant therapy approach in patients wi th rectal cancer. Surg Oncol Insights. 2024;1(2):100043.
20. Feng L, Liu Z, Li C, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in local ly advanced rectal cancer: a multicentre observational study. Lancet Digit Health. 202 2;4(1):e8-e17.
21. Shi R, Sun J, Zhou Z, et al. Deep multimodal fusion of patho-radiomic and clinic al data for enhanced survival prediction for colorectal cancer patients. NPJ Digit Med. 2025;8:221. doi:10.1038/s41746-025-02210-z.

7. FIGURE LEGENDS

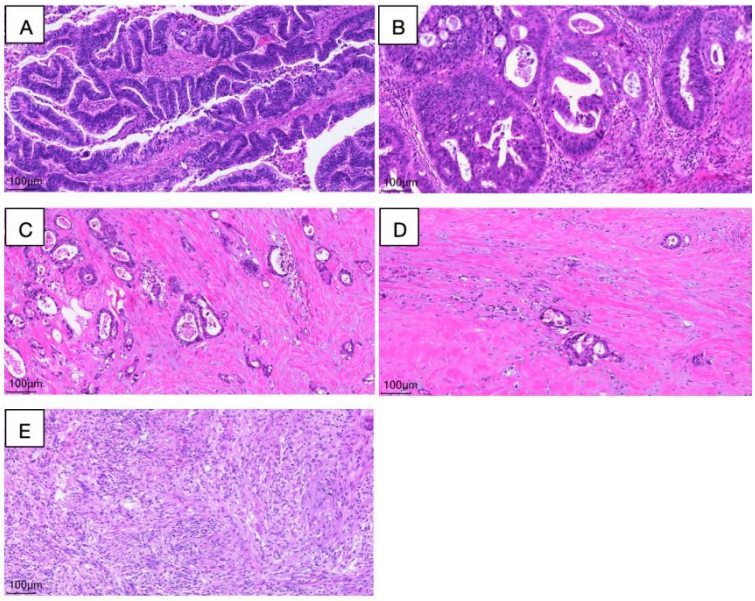


Figure 1. Tumor Regression Grade (TRG). A, B, C, D, and E corresp ond to TRG 4, TRG 3, TRG 2, TRG 1, and TRG 0, respectively.

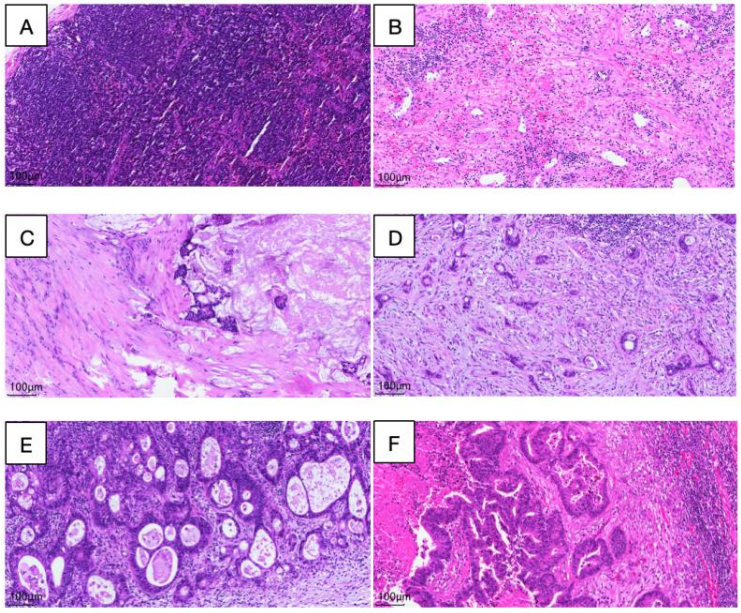


Figure 2. Lymph node regression grades (LRG). A, B, C, D, E, and F correspond to LRG 0, 1, 2, 3, 4, and 5, respectively.

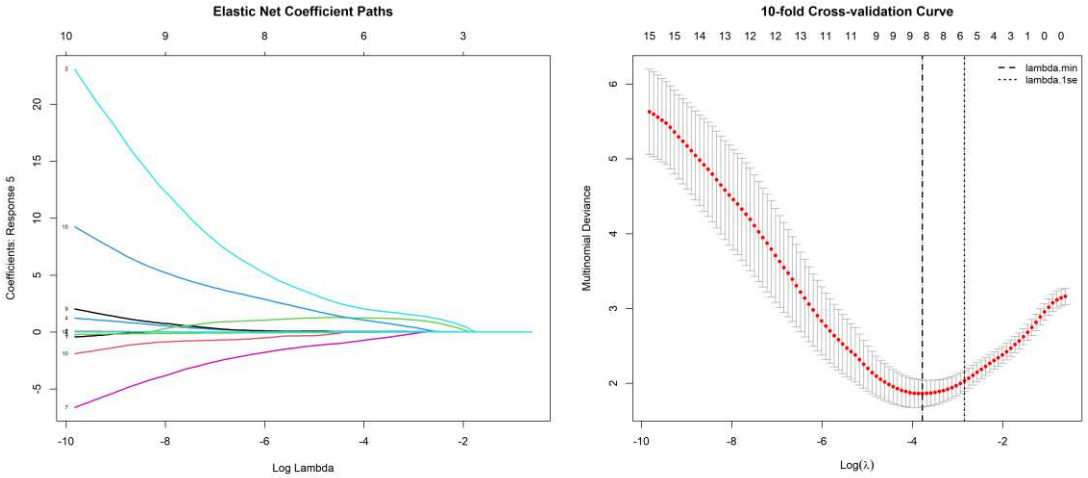


Figure 3. Elastic Net feature selection and cross-validation for optimal penalty parameter tuning. (A) Elastic Net coefficient paths. As the penalty parameter λ increases (moving from left to right on the log scale), the coefficients of less informative variables progressively shrink toward zero, retaining only key predictors strongly associated with LRGmax classification. (B) Ten-fold cross-validation curve. The optimal penalty parameter (λ_min) was selected at the value corresponding to the minimum multinomial deviance (dashed vertical line), balancing model parsimony and predictive accuracy. The dotted line indicates λ_1se, the largest λ within one standard error of the minimum.

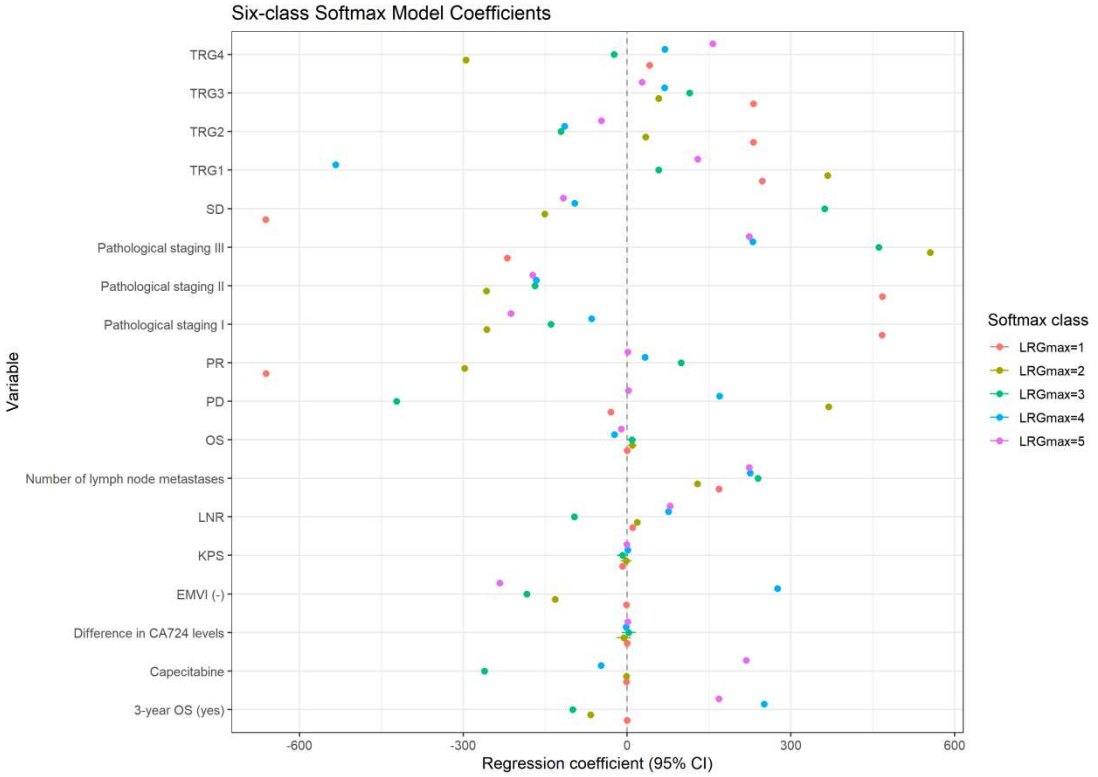


Figure 4. Elastic Net–Softmax multiclass classification model coefficient distribution plot (95% CI). The x-axis represents the regression coefficients (95%

CI), with 0 indicating the dashed reference line; the y-axis displays the 18 predictor variables. Color indicates the five Softmax categories (LRGmax = 1 to 5).

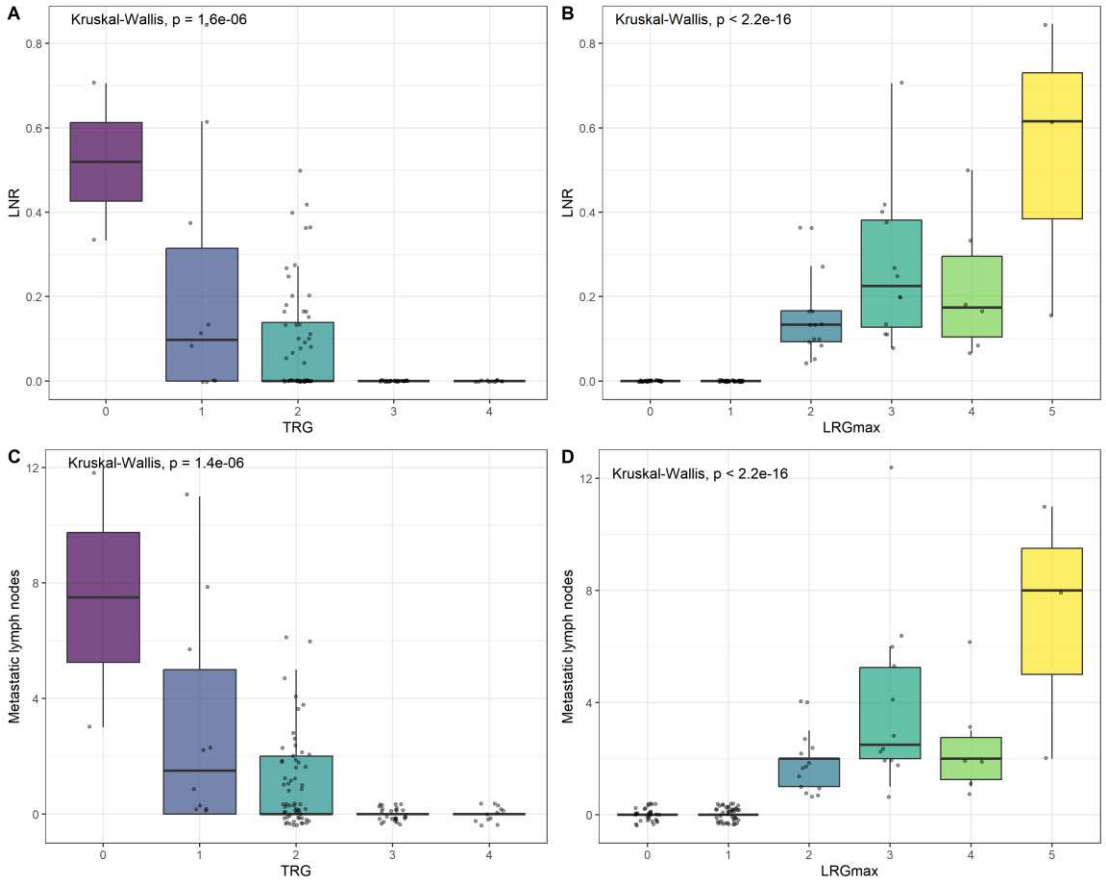


Figure 5. Association of tumor regression grade (TRG) and lymph node regression grade (LRGmax) with nodal metastatic burden. (A) Lymph node ratio (LNR) across TRG grades (0–4). (B) LNR across LRGmax grades (0–5). (C) Number of metastatic lymph nodes across TRG grades (0–4). (D) Number of metastatic lymph nodes across LRGmax grades (0–5).

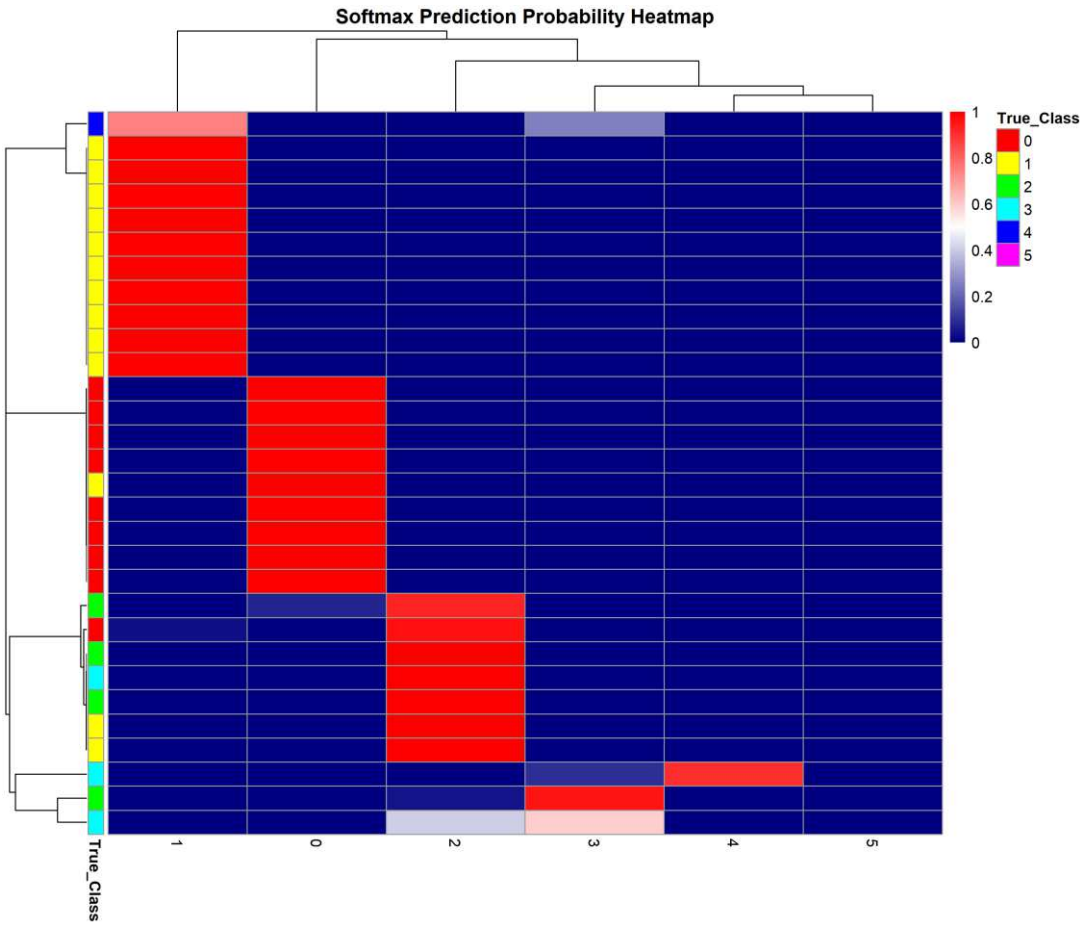


Figure 6. Softmax probability heatmap. The x-axis represents the predicted class (LRGmax); the y-axis represents the true class (LRGmax). Color intensity represents the probability value output by Softmax. Bright red indicates probability close to 1.0 (high confidence); dark blue indicates probability close to 0.0 (low confidence).

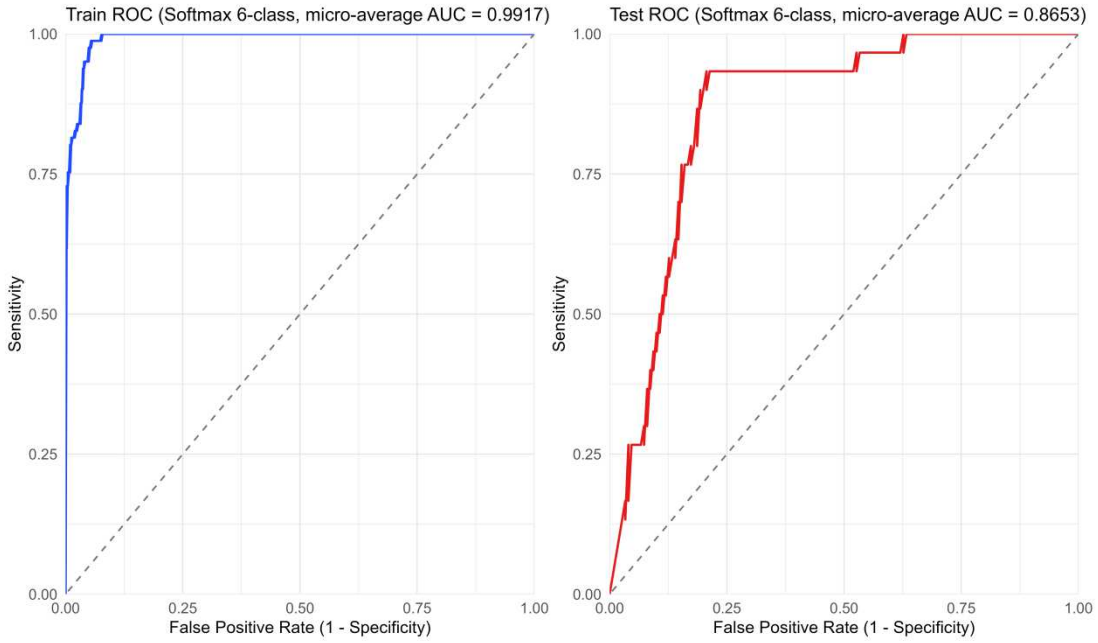


Figure 7. Receiver operating characteristic (ROC) curves of the Elastic Net–Softmax model in the training and validation sets. (Left) Training-set ROC curve (AUC = 0.9935). (Right) Validation-set ROC curve (AUC = 0.8653). The dashed diagonal line represents the reference line of no discrimination (AUC = 0.50).

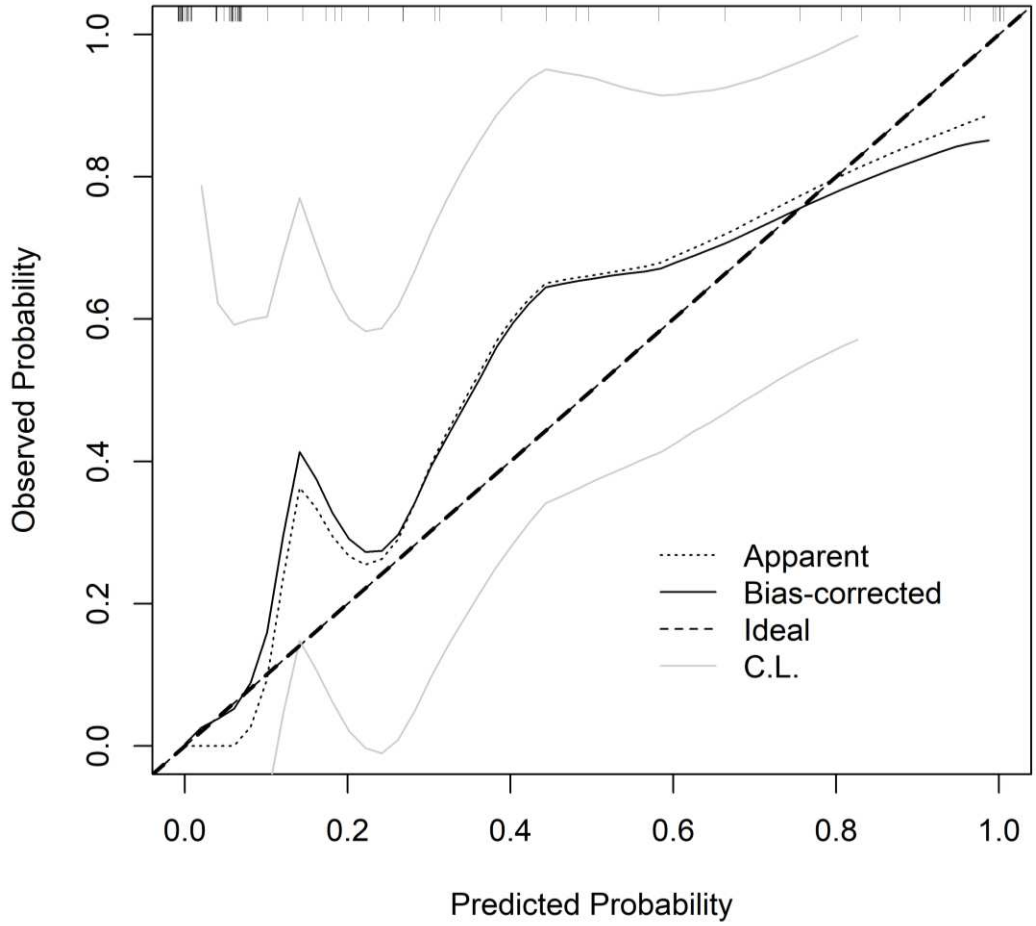


Figure 8. Calibration curve of the Elastic Net–Softmax model after bootstrap internal validation. The bias-corrected curve (solid line) closely overlaps with the ideal reference line (dashed diagonal), indicating high consistency between predicted probabilities and observed outcomes. The apparent curve (dotted line) represents the uncorrected training-set performance, and the confidence limits (gray lines) denote the 95% confidence intervals derived from 1,000 bootstrap resampling iterations.

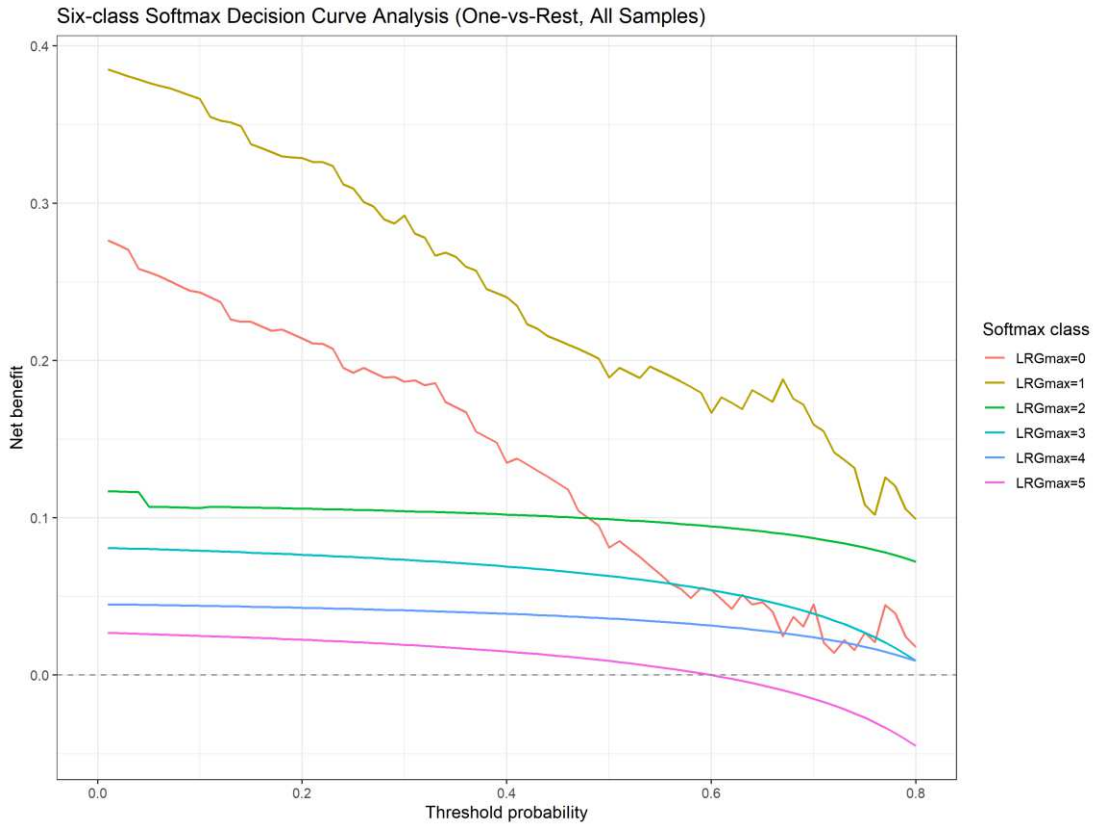


Figure 9. Decision curve analysis (DCA) evaluating the clinical net benefit of the Elastic Net–Softmax model.

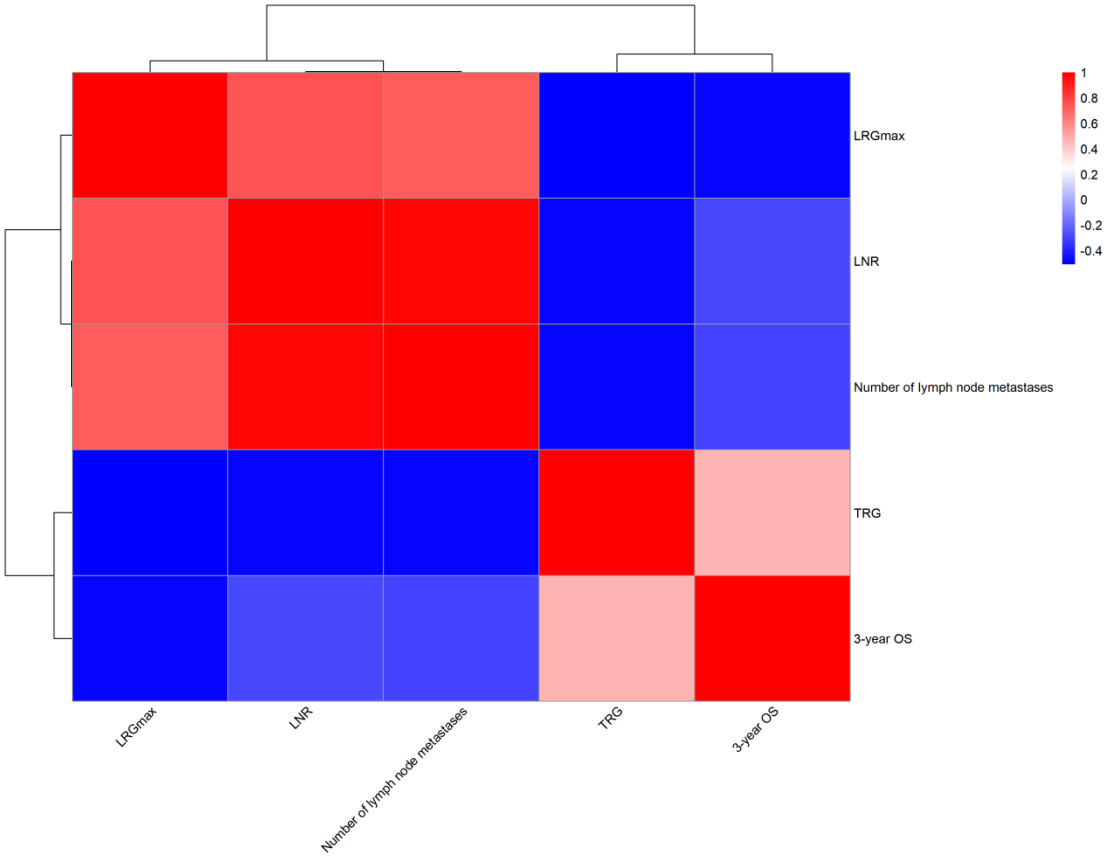


Figure 10. Correlation heatmap with hierarchical clustering of key clinicopathological and outcome variables. Spearman rank correlation coefficients were calculated between LRGmax, lymph node ratio (LNR), number of lymph node metastases, tumor regression grade (TRG), and 3-year overall survival (OS).

8. TABLES

Table 1. Clinical, pathological, and treatment characteristics of pathological LRGmax (0-5) response stratification in patients undergoing radical resection for rectal cancer to neoadjuvant chemoradiotherapy.

Variable	LRGmax0 (N=32)	LRGmax1 (N=44)	LRGmax2 (N=14)	LRGmax3 (N=12)	LRGmax4 (N=6)	LRGmax5 (N=3)	*p*
gender (%)							
male	20 (62.5)	33 (75.0)	8 (57.1)	8 (66.7)	4 (66.7)	2 (66.7)	0.823
female	12 (37.5)	11 (25.0)	6 (42.9)	4 (33.3)	2 (33.3)	1 (33.3)	
age	53.28	52.93	55.14	48.83	54.83	67.67	0.3
underlying diseases (%)							
no	14 (43.8)	26 (59.1)	7 (50.0)	8 (66.7)	4 (66.7)	1 (33.3)	0.611
yes	18 (56.2)	18 (40.9)	7 (50.0)	4 (33.3)	2 (33.3)	2 (66.7)	
Family history of genetic diseases (%)							
no	27 (84.4)	33 (75.0)	10 (71.4)	10 (83.3)	5 (83.3)	3 (100.0)	0.779
yes	5 (15.6)	11 (25.0)	4 (28.6)	2 (16.7)	1 (16.7)	0 (0.0)	
BMI	1.79	1.75	1.77	1.74	1.79	1.72	0.877
Tumor segmentation (%)							
upper	0 (0.0)	1 (2.3)	1 (7.1)	0 (0.0)	1 (16.7)	0 (0.0)	0.293
middle	8 (25.0)	16 (36.4)	4 (28.6)	7 (58.3)	1 (16.7)	1 (33.3)	
Lower	24 (75.0)	27 (61.4)	9 (64.3)	5 (41.7)	4 (66.7)	2 (66.7)	
Pathological differentiation degree (%)							
Poorly	3 (9.4)	6 (13.6)	2 (14.3)	4 (33.3)	2 (33.3)	0 (0.0)	0.597
Moderately	27 (84.4)	34 (77.3)	12 (85.7)	7 (58.3)	4 (66.7)	3 (100.0)	
Well	2 (6.2)	4 (9.1)	0 (0.0)	1 (8.3)	0 (0.0)	0 (0.0)	
cT (%)							0.304
2	2 (6.2)	2 (4.5)	2 (14.3)	1 (8.3)	0 (0.0)	0 (0.0)	

3	19 (59.4)	25 (56.8)	10 (71.4)	8 (66.7)	5 (83.3)	0 (0.0)	
4	11 (34.4)	17 (38.6)	2 (14.3)	3 (25.0)	1 (16.7)	3 (100.0)	
cN (%)							
0	7 (21.9)	8 (18.2)	2 (14.3)	0 (0.0)	0 (0.0)	0 (0.0)	
1	21 (65.6)	22 (50.0)	6 (42.9)	8 (66.7)	3 (50.0)	1 (33.3)	0.516
2	4 (12.5)	13 (29.5)	6 (42.9)	4 (33.3)	3 (50.0)	2 (66.7)	
3	0 (0.0)	1 (2.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Clinical staging(%)							
II	7 (21.9)	8 (18.2)	2 (14.3)	0 (0.0)	0 (0.0)	0 (0.0)	0.398
III	25 (78.1)	36 (81.8)	12 (85.7)	12 (100.0)	6 (100.0)	3 (100.0)	
MRF（rectal mesenteric fascia） (%)							
Positive	18 (56.2)	19 (43.2)	6 (42.9)	3 (25.0)	3 (50.0)	2 (66.7)	0.516
Negative	14 (43.8)	25 (56.8)	8 (57.1)	9 (75.0)	3 (50.0)	1 (33.3)	
EMVI（Invasion of extravascular vessels in the intestinal wall） (%)							
Positive	16 (50.0)	26 (59.1)	5 (35.7)	7 (58.3)	0 (0.0)	3 (100.0)	0.037
Negative	16 (50.0)	18 (40.9)	9 (64.3)	5 (41.7)	6 (100.0)	0 (0.0)	
Pre-treatment weight (kg)	68.76	67.39	69.07	67.67	72.50	67.33	0.941
Pre-treatment CEA(mmol/L)	11.72	15.41	10.61	10.87	10.75	12.48	0.961
Pre-treatment CA-724(mmol/L)	7.07	6.87	6.57	8.05	5.10	19.98	0.485
kps	90.31	89.09	92.86	90.00	90.00	90.00	0.007
Pre-treatment NRS)	0.50	0.59	0.43	0.50	0.33	1.33	0.251
chemotherapy regimen (%)							
Xelox	15 (46.9)	26 (59.1)	8 (57.1)	11 (91.7)	4 (66.7)	0 (0.0)	0.042
Capecitabine	17 (53.1)	18 (40.9)	6 (42.9)	1 (8.3)	2 (33.3)	3 (100.0)	

Bone marrow suppression response (%)							
0	11 (34.4)	17 (38.6)	5 (35.7)	4 (33.3)	1 (16.7)	2 (66.7)	
1	9 (28.1)	15 (34.1)	2 (14.3)	3 (25.0)	5 (83.3)	0 (0.0)	0.622
2	8 (25.0)	9 (20.5)	6 (42.9)	5 (41.7)	0 (0.0)	1 (33.3)	
3	3 (9.4)	2 (4.5)	1 (7.1)	0 (0.0)	0 (0.0)	0 (0.0)	
4	1 (3.1)	1 (2.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Radiation proctitis (%)							
0	13 (40.6)	19 (43.2)	8 (57.1)	3 (25.0)	2 (33.3)	2 (66.7)	
1	15 (46.9)	16 (36.4)	4 (28.6)	8 (66.7)	1 (16.7)	1 (33.3)	0.14
2	4 (12.5)	9 (20.5)	1 (7.1)	1 (8.3)	2 (33.3)	0 (0.0)	
3	0 (0.0)	0 (0.0)	1 (7.1)	0 (0.0)	1 (16.7)	0 (0.0)	
radiation dermatitis (%)							
0	25 (78.1)	34 (77.3)	13 (92.9)	8 (66.7)	5 (83.3)	3 (100.0)	
1	7 (21.9)	8 (18.2)	1 (7.1)	4 (33.3)	0 (0.0)	0 (0.0)	0.439
2	0 (0.0)	1 (2.3)	0 (0.0)	0 (0.0)	1 (16.7)	0 (0.0)	
3	0 (0.0)	1 (2.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Therapeutic effect evaluation (%)							
PD	0 (0.0)	1 (2.3)	3 (21.4)	0 (0.0)	1 (16.7)	0 (0.0)	
SD	3 (9.4)	1 (2.3)	5 (35.7)	4 (33.3)	4 (66.7)	2 (66.7)	<0.001
PR	26 (81.2)	36 (81.8)	6 (42.9)	8 (66.7)	1 (16.7)	1 (33.3)	
CR	3 (9.4)	6 (13.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
CEA after treatment(mmol/L)	5.32	7.02	8.79	3.17	3.37	2.11	0.909
Difference in CEA levels before and after treatment (mmol/L))	-6.40	-8.39	-1.82	-7.70	-7.38	-10.37	0.963

CA-724 after treatment (mmol/L)	3.03	5.66	19.80	2.10	7.08	1.31	0.095
Difference in CA-724 levels before and after treatment (mmol/L))	-4.04	-1.21	13.22	-5.95	1.99	-18.68	0.027
Post-treatment weight (mmol/L)	69.36	67.39	67.00	67.38	71.67	64.67	0.925
Difference in weight change before and after treatment(kg)	0.60)	0.00	-2.07	-0.29	-0.83	-2.67	0.471
Interval between surgeries (week)	11.09	9.68	10.86	10.25	10.00	9.33	0.515
Surgical methods(%)							
Dixon	28 (87.5)	30 (68.2)	12 (85.7)	11 (91.7)	3 (50.0)	3 (100.0)	0.086
Miles	4 (12.5)	14 (31.8)	2 (14.3)	1 (8.3)	3 (50.0)	0 (0.0)	
Number of lymph nodes removed	13.59	14.50	13.36	14.25	11.83	13.00	0.433
Number of lymph node metastases	0.00	0.00	1.93)	3.92	2.50	7.00	<0.001

							<0.00
LNR	0.00	0.00	0.16	0.27	0.22	0.54	1
pT (%)							
0	6 (18.8)	6 (13.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
1	2 (6.2)	0 (0.0)	1 (7.1)	0 (0.0)	0 (0.0)	0 (0.0)	0.029
2	8 (25.0)	14 (31.8)	2 (14.3)	1 (8.3)	0 (0.0)	0 (0.0)	
3	13 (40.6)	19 (43.2)	11 (78.6)	10 (83.3)	4 (66.7)	1 (33.3)	
4	3 (9.4)	5 (11.4)	0 (0.0)	1 (8.3)	2 (33.3)	2 (66.7)	
pN (%)							
0	32 (100.0)	44 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
1	0 (0.0)	0 (0.0)	12 (85.7)	7 (58.3)	5 (83.3)	1 (33.3)	<0.00
2	0 (0.0)	0 (0.0)	2 (14.3)	4 (33.3)	1 (16.7)	0 (0.0)	1
3	0 (0.0)	0 (0.0)	0 (0.0)	1 (8.3)	0 (0.0)	2 (66.7)	
TRG (%)							
0	0 (0.0)	0 (0.0)	0 (0.0)	1 (8.3)	1 (16.7)	0 (0.0)	
1	0 (0.0)	4 (9.1)	1 (7.1)	3 (25.0)	0 (0.0)	2 (66.7)	<0.00
2	13 (40.6)	20 (45.5)	13 (92.9)	8 (66.7)	5 (83.3)	1 (33.3)	1
3	13 (40.6)	14 (31.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
4	6 (18.8)	6 (13.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Pathological staging (%)							
0	6 (18.8)	6 (13.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
I	8 (25.0)	10 (22.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	<0.00
II	18 (56.2)	28 (63.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1
III	0 (0.0)	0 (0.0)	14 (100.0)	12 (100.0)	6 (100.0)	3 (100.0)	
Local recurrence (%)							
no	27 (84.4)	41 (93.2)	12 (85.7)	12 (100.0)	4 (66.7)	3 (100.0)	0.248
yes	5 (15.6)	3 (6.8)	2 (14.3)	0 (0.0)	2 (33.3)	0 (0.0)	
Distant transfer (%)							
no	27 (84.4)	39 (88.6)	12 (85.7)	9 (75.0)	5 (83.3)	3 (100.0)	0.851

yes	5 (15.6)	5 (11.4)	2 (14.3)	3 (25.0)	1 (16.7)	0 (0.0)	
Overall Survival（OS） (mean (SD))	56.97 (19.58)	35.66 (14.15)	51.21 (28.21)	31.50 (15.17)	23.67 (8.16)	8.33 (4.04)	<0.001
3-year OS (%)							
no	2 (6.2)	19 (43.2)	4 (28.6)	9 (75.0)	5 (83.3)	3 (100.0)	<0.001
yes	30 (93.8)	25 (56.8)	10 (71.4)	3 (25.0)	1 (16.7)	0 (0.0)	

Table 2. Characteristics of patients in the training and validation sets.

Variable		training set(N=79)	Validation set(N=32)	*p*
gender (%)				
	male	56 (70.9)	19 (59.4)	0.342
	female	23 (29.1)	13 (40.6)	
age (mean (SD))		53.30 (12.30)	53.53 (12.60)	0.93
underlying diseases (%)				
	no	46 (58.2)	14 (43.8)	0.24
	yes	33 (41.8)	18 (56.2)	
Family history of genetic diseases (%)				
	no	64 (81.0)	24 (75.0)	0.653
	yes	15 (19.0)	8 (25.0)	
Pre-treatment weight (mean (SD))		68.01(11.7)	69.03 (10.73)	0.669
Body surface area (mean (SD))		1.77 (0.17)	1.76 (0.14)	0.858
Tumor segmentation (%)				
	upper section	1 (1.3)	2 (6.2)	0.29
	middle section	28 (35.4)	9 (28.1)	
	Lower section	50 (63.3)	21 (65.6)	

Variable	training set(N=79)	Validation set(N=32)	*p*
Pathological differentiation degree (%)			
Poorly differentiated adenocarcinoma	13 (16.5)	4 (12.5)	
Moderately differentiated adenocarcinoma	61 (77.2)	26 (81.2)	0.869
Well-differentiated adenocarcinoma	5 (6.3)	2 (6.2)	
cT (%)			
2	4 (5.1)	3 (9.4)	
3	48 (60.8)	19 (59.4)	0.692
4	27 (34.2)	10 (31.2)	
cN (%)			
0	10 (12.7)	7 (21.9)	
1	42 (53.2)	19 (59.4)	0.332
2	26 (32.9)	6 (18.8)	
3	1 (1.3)	0 (0.0)	
Clinical staging (%)			
II	10 (12.7)	7 (21.9)	0.352
III	69 (87.3)	25 (78.1)	
MRF (%)			
Positive	37 (46.8)	14 (43.8)	0.932
Negative	42 (53.2)	18 (56.2)	
EMVI (%)			
Positive	43 (54.4)	14 (43.8)	0.418
Negative	36 (45.6)	18 (56.2)	
Pre-treatment CEA(mean (SD))	13.68 (23.39)	11.02 (14.68)	0.552

Variable	training set(N=79)	Validation set(N=32)	*p*
Pre-treatment CA-724 (mean (SD))	6.63 (11.32)	8.87 (9.75)	0.33
kps (mean (SD))	90.38 (3.38)	89.38 (2.46)	0.13
Pre-treatment NRS (mean (SD))	0.54 (0.64)	0.53 (0.57)	0.92
chemotherapy regimen (%)			
Xelox	45 (57.0)	19 (59.4)	0.983
Capecitabine	34 (43.0)	13 (40.6)	
Bone marrow suppression response (%)			
0	29 (36.7)	11 (34.4)	
1	26 (32.9)	8 (25.0)	0.234
2	20 (25.3)	9 (28.1)	
3	2 (2.5)	4 (12.5)	
4	2 (2.5)	0 (0.0)	
Radiation proctitis (%)			
0	32 (40.5)	15 (46.9)	
1	35 (44.3)	10 (31.2)	0.106
2	12 (15.2)	5 (15.6)	
3	0 (0.0)	2 (6.2)	
radiation dermatitis (%)			
0	65 (82.3)	23 (71.9)	
1	14 (17.7)	6 (18.8)	0.052
2	0 (0.0)	2 (6.2)	
3	0 (0.0)	1 (3.1)	
Therapeutic effect evaluation (%)			
PD	3 (3.8)	2 (6.2)	0.087
SD	17 (21.5)	2 (6.2)	
PR	55 (69.6)	23 (71.9)	

Variable	training set(N=79)	Validation set(N=32)	*p*
CR	4 (5.1)	5 (15.6)	
CEA after treatment(mean (SD))	6.40 (16.37)	5.03 (10.91)	0.664
Difference in CEA levels before and after treatment (mean (SD))	-7.28 (26.10)	-6.00 (8.51)	0.786
CA-724 after treatment (mean (SD))	6.30 (19.31)	6.17 (16.92)	0.973
Difference in CA-724 levels before and after treatment (SD))	-0.33 (19.87)	-2.70 (16.67)	0.554
Post-treatment weight (mean (SD))	68.18 (12.34)	67.77 (11.51)	0.869
Difference in weight change before and after treatment(mean (SD))	0.18 (4.95)	-1.27 (2.70)	0.123
Interval between surgeries (mean (SD))	9.92 (3.06)	11.25 (3.46)	0.049
Surgical methods (%)			
Dixon	61 (77.2)	26 (81.2)	0.831
Miles	18 (22.8)	6 (18.8)	
Number of lymph nodes removed (mean (SD))	13.66 (2.83)	14.44 (4.17)	0.257
Number of lymph node metastases (mean (SD))	1.03 (2.29)	0.91 (1.61)	0.789
LNR (mean (SD))	0.08 (0.16)	0.07 (0.13)	0.965
pT (%)			
0	6 (7.6)	6 (18.8)	0.288
1	3 (3.8)	0 (0.0)	
2	18 (22.8)	7 (21.9)	

Variable	training set(N=79)	Validation set(N=32)	*p*
3	41 (51.9)	17 (53.1)	
4	11 (13.9)	2 (6.2)	
pN (%)			
0	54 (68.4)	22 (68.8)	
1	18 (22.8)	7 (21.9)	0.594
2	4 (5.1)	3 (9.4)	
3	3 (3.8)	0 (0.0)	
TRG (%)			
0	1 (1.3)	1 (3.1)	
1	9 (11.4)	1 (3.1)	0.297
2	43 (54.4)	17 (53.1)	
3	20 (25.3)	7 (21.9)	
4	6 (7.6)	6 (18.8)	
Pathological staging (%)			
0	6 (7.6)	6 (18.8)	
I	13 (16.5)	5 (15.6)	0.366
II	35 (44.3)	11 (34.4)	
III	25 (31.6)	10 (31.2)	
Local recurrence (%)			
no	72 (91.1)	27 (84.4)	0.483
yes	7 (8.9)	5 (15.6)	
Distant transfer (%)			
no	68 (86.1)	27 (84.4)	1
yes	11 (13.9)	5 (15.6)	
OS (mean (SD))	40.80 (21.10)	44.72 (22.37)	0.385
3-year OS (%)			0.487

Variable	training set(N=79)	Validation set(N=32)	*p*
no	32 (40.5)	10 (31.2)	
yes	47 (59.5)	22 (68.8)	

Table 3. Model performance evaluation results.

Data set	Accuracy	AUC	Precision	Recall	Sensitivity	Specificity
Train	0.88	0.99	0.94	0.94	0.94	0.97
Test	0.43	0.86	0.30	0.30	0.30	0.87

Table 4. Correlation of tumor regression grade (TRG) and lymph node regression grade (LRGmax) with clinical outcomes.

Analysis	Method	Statistic	P-value
TRG vs. radiotherapy efficacy	Chi-square	82.03	< 0.001
LRGmax vs. radiotherapy efficacy	Spearman *r*	0.28	0.003
TRG vs. distant metastasis	Chi-square	4.54	0.338
LRGmax vs. distant metastasis	Spearman *r*	0.02	0.854
TRG vs. 3-year overall survival	Chi-square	26.15	< 0.001
LRGmax vs. 3-year overall survival	Spearman *r*	−0.47	< 0.001

9. DECLARATIONS

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Affiliated Cancer Hospital of Xinjiang Medical University (approval number: K-2025176). Written informed consent was obtained from all patients. The study was conducted in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This project was funded by the Natural Science Youth Research Project of Xinjiang Medical University.（Project No. 2024XYZR34）

Authors' contributions

first author：Kai Liu,Conceptualization, Methodology, Formal analysis, Software, Visualization, Writing – original draft.
co-first author：Shuyuan Zhao,Data curation, Investigation, Validation, Writing – review & editing.
second author：Ran Zhang， Pathological Diagnosis and TRG/LRG Grading in Rectal Cancer Patients， Resources, Project administration, Writing – review & editing.
Corresponding author：Yan Liu,Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.docx